

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Emotional Appraisal of Moral Dilemma in Characters

### This is the author's manuscript

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/129620> since 2016-11-30T15:01:15Z

*Publisher:*

Springer

*Published version:*

DOI:10.1007/978-3-642-34851-8\_15

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

# Emotional appraisal of moral dilemma in characters

Cristina Battaglino<sup>1</sup> and Rossana Damiano<sup>1</sup>

Università di Torino, Dipartimento di Informatica and CIRMA  
battagli@di.unito.it rossana@di.unito.it

**Abstract.** In the narrative realm, emotions in general, and moral emotions in particular, play an important role in conveying moral values to the audience. A paradigmatic case is given by the occurrence of moral dilemmas in stories, where the character must trade off options that cannot be reconciled, each related with a certain moral value. In this paper, we propose a model of the appraisal of moral emotions, and describe an agent architecture that reacts to the occurrence of a moral dilemma with the appropriate emotions.

**Keywords:** Virtual characters, Emotions, Moral values

## 1 Introduction

Moral values have been acknowledged by drama theorists and psychologists as having a primary role in drama. According to Bruner, “stories achieve their meanings by explicating deviations from the ordinary in a comprehensible form” [6], contributing to enforce the canonicity of a culture’s values. In scriptwriting, the moral nature of the narrative premise, pointed out since the pioneering work of [12, 22], becomes explicit in the work by [28]. Moral dilemmas take to the stage the representation of diverging moral commitments within the same character, bringing them to life with the help of the appropriate emotional range. In moral dilemma, a character is faced with options that cannot be reconciled: the choice of one option determines the impossibility of bringing about the others, and each corresponds to a different moral value, so that the character is faced with a choice between different values. This deadlock makes contrasting emotions arise in the character, deeply interwoven with rational considerations. Finally, an option (and a value) prevails, but the emergence of canonicity does not retain the character from experiencing strong and conflicting emotional states along the process.

In this paper, we propose a character model in which moral emotions become part of the deliberation process of the character. Moral emotions, like pride and shame, arise as a consequence of moral values; their intensity depends on the importance of the values they are related with and is increased if these values are shared with other agents. The aim is to create artificial characters who display the emotional range triggered by the occurrence of a moral dilemma, defined

as mutually exclusive options that put different values at stake. In the struggle among conflicting values, the choice of an option by the character is driven by the intensity of the moral emotions she experiences; a special role is played by the values shared with the community, which increase the intensity of shame and pride. Remorse is also part of the game, when the character, after deliberating and acting, is left alone with her conscience.

This paper is structured as follows. After surveying the related work (Section 2), in Section 3 we describe the relationship between moral emotions and values and the agent model. Section 4 illustrates the agent architecture where moral emotions are embedded. Section 5 shows the model at work on a literary example (the story of “Iphigenia in Aulis”, written by Euripides). Conclusions and future work end the paper.

## 2 Related work

The construction of a model of character who reacts to moral dilemma in the same way as literary characters requires the integration of a rational and an emotional component. Both components should embed some notion of moral value, to let the character reason about the compliance with values at the behavioral level and feel moral emotions when her values are put at stake.

Intelligent agents offer an operational way to design and implement characters in interactive storytelling, as shown by a number of successful applications [20, 23]. The Belief-Desire-Intention (BDI) model of agent [5] has proven to be an effective basis for the implementation of artificial characters, thanks to the availability of programmable agent frameworks [21, 4]. In contemporary aesthetics, the adequacy of this model to the paradigm of identification postulated by [8] has been put forth by the work of Feagin [15].

Many works tried to integrate computational models of emotions in a cognitive architecture for intelligent agents [24, 14, 18]. Although different theories of emotions have been proposed (including physiological and dimensional models), most computational models are based on appraisal theory, in which cognitive processes are involved in the generation of emotions [19, 17, 25]. According to appraisal theories, cognitive processes have the function of building a mental representation of the situation in which a person is involved. This representation (*person-environment relation*) is not limited to the external environment, but includes also the internal disposition of a person as goals, desires, intentions, norms, moral rules. Emotions arise from appraisal of the person-environment relation according to appraisal dimensions that are defined in the theory (i.e. desirability of an event). Many models don’t take into account the link between moral emotions, values and goals [16, 18, 14, 3], with the relevant exception of FLAME [13]. In FLAME, moral rules are acquired by learning user’s actions and a fuzzy logic approach is used to calculate the intensity of emotions; however, the appraisal process is not encoded in the system.

In the OCC theory [19], the person-environment relation is represented by goals, standards and attitudes; appraisal dimensions are represented by *desir-*

*ability* (or *undesirability*) of an event, *praiseworthiness* (or *blameworthiness*) of an action, *liking* (or *disliking*) of an object. According to this model, the agent’s “standards” (i.e. the agent’s “beliefs in term of which moral and other kinds of judgmental evaluations are made”) affect the evaluation of self and others’ actions. Actions that meet the agent’s standards are deemed praiseworthy, and their execution triggers emotions like pride and admiration. Conversely, blameworthy actions trigger emotions like shame and reproach. So, in our work, we assume the OCC model, and rely on previous work by [7] to establish an explicit link between moral values and moral emotions.

In the story generation systems by [2], a finite set of dilemmas are defined, each characterised by a set of preconditions that the system has the task to achieve. Generated dilemmas are presented to the user in order to engage her in the drama with the choice of the next course of action. Differently from [2], we present a model of character in which a moral conflict arises from the conflict of values and from the emotional appraisal of one’s own actions. Inspired by work on moral dilemma [2] and the notion of value [11, 26], we propose a model of how a character reacts in front of a moral dilemma and how moral emotions can affect her behavior. Different individuals acknowledge different values, arranged into subjective scales [26]. So, different characters react differently to values at stake, as a consequence of the values they care for and the importance they attribute to them.

### 3 Dilemma-compliant agent model

In previous work [10], value-sensitive agents are modeled as BDI agents, augmented with the notion of value. An agent features a set of values, arranged into a subjective ‘scale of values’ [26]. Each value is associated with a set of conditions: when one or more conditions hold in the state of the world, the value is put at stake. The value-sensitive agent monitors the state of the world for values at stake: when the agent realizes that some value is at stake, it modifies its commitment accordingly, by forming a goal (value-dependent goal) that contributes to re-establishing the value (or the values) at stake. Notice that, according to this model, the monitoring of values is carried out not only on what the agent believes to be the current state of the world but also on the agent’s expectations about the outcomes of the events and of the other agent’s actions.

Following the model described in [7], we map the notion of agent’s “standards” on the notion of “values” and we define the *praiseworthiness* or *blameworthiness* of an action on the basis of the agent’s values. The agent considers her own action as praiseworthy only if she has formed the intention to execute it to bring a value at stake back to balance. Conversely, if an action puts at stake a value of the appraising agent, it is considered blameworthy. The role of values is relevant not only for the appraisal of an agent’s own actions, but also for the appraisal of other agents’ behavior. In our model, different emotions can arise in the agent, and an action can be appraised as blameworthy and praiseworthy at the same time. Also, we distinguish when a value is put at stake by an action

but can be reestablished subsequently and when a value is put at stake with no possibility of being reestablished. Based on this, in the following section, we introduce moral dilemmas.

We model a character as a BDI agent augmented with the notion of values and emotions. So, a character is a 5-tuple  $\{B, D, I, V, E\}$  where  $B$  is the set of beliefs of the character,  $D$  is the set of desires (or high-level goals),  $I$  is the set of intentions,  $V$  is the set of values,  $E$  is the set of emotions. For the aim of this work, we consider the emotions of pride and shame, classified in OCC model as Attribution emotions concerning the self. The beliefs base includes not only the state of the world but also expectations about how it may evolve, information about interpersonal relationships with others agent and their mental state (clearly, in a real implementation a theory of mind about others agents and expectations have to be restricted to a certain number of nesting).

Goals arise as a consequence of the character's values. Inspired by [27], a goal  $g$  is defined by a tuple  $g(c, e, s, \pi, t)$  where:

- $c$  is the satisfying condition. When  $c$  is true in the state of the world the goal is achieved and dropped;
- $e$  is the failure condition. When  $e$  is true in the current state of the world, the goal is dropped;
- $s$  is the state of the goal. A goal can be adopted, suspended, activated or dropped. Following [27], desires  $\in D$  are adopted goals. Adopted goals remain suspended until they are ready for execution (i.e., they are in active state), then possibly suspended again if a more important goal is adopted. Goals can eventually be dropped if certain conditions hold, namely, when the rationality constraints stated by [9] are met;
- $\pi$  is the plan formed by the planning component to achieve the goal;
- $t$  is the goal type. In this work, we consider only achievement goals and perform goals. When dealing with values, an achievement goal is the goal to achieve a certain state of affairs in which the value is not at stake anymore, while a perform goal is the goal to execute actions (i.e. plans) to re-establish the value at stake.

A value  $v$  is defined by a set of constructs of the form  $v(c, r, l)$  where:

- $c$  is a ground formula. When  $c$  holds in the state of the world or in the character's expectation the value is put at stake;
- $r$  (a real number) is the priority of the value. Values are organized in a scale of values [26];
- $l$  represents the probability of reestablishing a value through the execution of a plan (this implies that the value is at stake and that the agent has an active value-dependent goal with a formed plan).

With respect to [27], [11], we introduce for plans a structure that contains information about the values at stake, in order to let the agent reason about dilemmas. A plan structure  $\pi \in I$  contains a sequence of actions, devised by the planning

component. We assume that actions are in STRIPS-like style with preconditions and effects. A plan  $\pi$  is described by a tuple  $(P, V^T, V^B, g, u)$  where:

- $P$  is the set of facts that the plan  $\pi$  makes true in the state of the world (if completely executed with success). Every condition  $p_i \in P$  is a condition  $c$  of a value at stake  $v_i \in V^T$  that may hold in the current state or in the future, according to agent's prospect reasoning (or expectations);
- $V^T$  is the set of values  $v_i$  put at stake by the plan  $\pi$ ;
- $V^B$  is the set of values  $v_i$  reestablished by the plan  $\pi$ ;
- $g$  is the value-dependent goal achieved by the plan. Goal  $g$  arises from the motivation of bringing values in  $V^B$  back to balance;
- $u$  is the associate utility (a real number). Agents choose which plans to execute with respect to their utility, calculated by taking into consideration the moral emotions that arise from reestablishing or putting at stake a value and the probability of success of plan.

Emotions are defined by  $(t, i, v, \pi)$  where:

- $t$  is the emotion type, according to OCC model;
- $i$  is the intensity of emotion (a real number); it is involved in the calculation of plan utility;
- $v$  is the emotion valence that indicates if emotions is a positive ('+') or negative ('-') emotion. For example, a positive emotions, like 'pride', contributes in positive way to utility and expected utility of a plan.
- $\pi$  is the plan from which emotion rises.

Further, in the belief base, we also define the following data structures:

- $V_{atStake}$ : the set of values  $v_i \in V$  put at stake in the current state of the world or in character's expectations;
- $V_{Shared}$ : the set of values shared with others agents, augmented with information about the relationship of the agent with them. A shared value  $sv_i \in V_{Shared}$  is described by  $(ag, v_i, rel(ag))$  where  $ag$  is the agent that shares the value,  $v_i$  is the shared value and  $rel(ag)$  is the believed intensity of the relationship with that agent.

## 4 Character's architecture

In order to cast the model described in the previous section into an agent architecture, we define the following agent loop:

**Monitoring:** the agent perceives the world and updates its beliefs ( $update(B)$ ). In particular, the agent supervises the failure condition on his/her goals to drop them and the conditions of her values at stake to control if a value at stake is re-established ( $checkConditions(D, V_{atStake})$ ).

**Goal adoption:** the agent updates its value at stake.  $\forall v_i \in V$  if the condition  $c$  of the value  $v_i$  holds in the state of the world, the value is at stake and

---

**Algorithm 1** Moral dilemma reappraisal
 

---

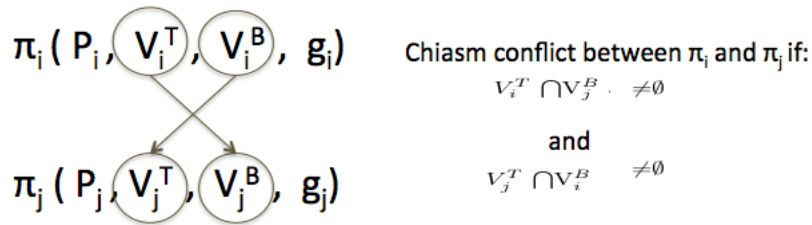
```

while true do
   $B \leftarrow \text{update}(B)$ 
   $\text{checkConditions}(D, V_{atStake})$ 
   $V_{atStake} \leftarrow \text{update}(V_{atStake})$ 
   $\text{formGoal}(D, V_{atStake})$ 
   $\text{generatePlans}(D, I)$ 
   $\text{prospectReasoning}(I, D, V)$ 
   $\text{checkConflict}(I)$ 
   $\text{selfAppraisal}(E, I, D, V_{atStake})$ 
   $\text{socialAppraisal}(E, I, D, V_{atStake}, V_{Shared})$ 
   $\pi \leftarrow \text{chooseBestPlan}(I)$ 
   $\text{execute}(\pi)$ 
   $\text{generateEmotions}(\pi)$ 
end while
  
```

---

is added to  $V_{atStake}$  ( $\text{update}(V_{atStake})$ ). The agent forms and adopts a value-dependent goal to restore the values at stake ( $\text{formGoals}(D, V_{atStake})$ ). For each new generated plans, the possibility of success is calculated to bring the value back to balance.

**Option Generation:** the agent performs means-end reasoning to generate new plans for the adopted goals ( $\text{generatePlans}(D, I)$ ). Then, she performs anticipatory reasoning to detect if the new plans put at stake some other value in  $V$  ( $\text{prospectReasoning}(I, D, V)$ ).  $\forall \pi_i \in I$ , the agent checks if an effect of an action puts another value  $v_i \in V$  at stake. Although expensive, focused anticipatory reasoning is an essential element of social frameworks, such as [18, 1]. If so, the condition of the value at stake is added to the plan set of precondition  $P$  and the value at stake is added to the character set  $V_{atStake}$  and the plan set  $V^T$ . Also, the agent forms a value-dependent goal to re-establish the new values at stake and generates new plans for new adopted goals. For every new plan, the probability of success is calculated to set the probability to re-establishing the value (the field  $l$  in value construct). The agent performs a check on the updated



**Fig. 1.** Definition of chiasmus between options.

set  $I$  to detects conflicts between plans ( $checkConflict(I)$ ). Plans are in conflict when a chiasmus is established between them. A chiasmus (Fig. 1) between two plans  $\pi_i$  and  $\pi_j$  exists when the intersections  $V_i^T \cap V_j^B$  and  $V_j^T \cap V_i^B \neq \emptyset$ . A conflict between a plan  $\pi_i$  and a plan  $\pi_j$  means that, when successfully executed  $\pi_j$ , the probability of success of plan  $\pi_i$  is equal to zero, and vice versa. So, the probability  $l$  of a value is set to the conditional probability of success of the related plan and not to the probability of success of the plan (as described above).

Making this assumption, we model the fact that there is no possibility of reestablishing the value in the future. So a moral dilemma is detected, the agent's decision is not about what value to bring in a positive state before, but which value must be put aside.

**Appraisal-based deliberation:** in this phase, the agent decides what goal becomes activated. The appraisal takes into consideration moral emotions to choose the value-dependent goal to activate and forms the value-dependent perform goal to execute the related plan. The deliberation is done in two phases: one is the self-appraisal ( $selfAppraisal(E, I, D, V_{atStake})$ ) and the other is the social appraisal ( $socialAppraisal(E, I, D, V_{atStake}, V_{Shared})$ ) of plans. In the self-appraisal phase the character considers only his own goals, while, in the social-appraisal phase, the values shared with the society are taken into account in the decision process.

In *self-appraisal*, the intensity of emotions that arise from a certain plan are calculated and the expected utility of the plan is computed:

1.  $Pride_I = \sum_{v_i \in V^B} r_i * \text{step}(\pi_i) * \text{probOfSucc}(\pi_i)$
2.  $Shame_I = \sum_{v_i \in V^T} r_i * (1 - l_i) * \text{step}(\pi_i) * \text{probOfSucc}(\pi_i)$
3.  $\text{expected-utility}(\pi_i) = Pride_I - Shame_I$

After the self-appraisal, the agent executes *social-appraisal*, in which she takes into account her beliefs about shared values and social relations to calculate the intensity of emotions and the expected utility of plans. So, in this phase, a plan is appraised as praiseworthy or blameworthy by taking into consideration the shared values. The priority of the shared value is obtained from the weighted sum between the priority of the shared value and the intensity of the relation with the agent with who shares that value. The new priorities of the values  $\in V^T$  or  $\in V^B$  are re-calculated as the sum of original priority with the priority of the shared value (for example, given  $v_i \in V^B$  and  $v_{i1}, v_{i2} \in V_{Shared}$  the new priority is equal to  $r_i \in V^B + \sum_{v_{ij} \in V_{Shared}} r_{ij}$ ). The intensity of emotions are then calculated by taking into consideration the new priority of the values, according to the formulas:

1.  $\text{shared-priority}(v_i) = \sum_{v_i \in V_{Shared}} r_i * \text{rel}(ag)$  given a value  $v_i \in V^T$  or  $V^B$ ;
2.  $Pride_I = (\sum_{v_i^B \in V^B} r_i + \text{shared-priority}(v_i)) * \text{step}(\pi_i) * \text{probOfSucc}(\pi_i)$
3.  $Shame_I = (\sum_{v_i^T \in V^T} r_i * (1 - l_i) + \text{shared-priority}(v_i)) * \text{step}(\pi_i) * \text{probOfSucc}(\pi_i)$



$$4. \text{ expected-utility}(\pi_i) = \text{Pride}_I - \text{Shame}_I$$

The agent chooses the best plan ( $\text{chooseBestPlan}(I)$ ) and forms a perform goal to execute the plan. The goal related to the chosen plans become activated, the others goals remain suspended.

**Execution** The agent executes the next action of activated plan – the first action when the plan is activated for the first time ( $\text{execution}(\pi)$ ,  $\text{generateEmotions}(\pi)$ ).

## 5 Example

The example, taken from the greek tragedy “Iphigenia in Aulis” by Euripides, revolves around Agamemnon, the leader of the Greek fleet during the Trojan War, and his decision to sacrifice his daughter, Iphigenia. We model the moment in which the Greek fleet is waiting at Aulis (Boeotia), but is unable to set sail due to a lack of wind. After consulting the seer Calchas, Agamemnon comes to know that the goddess Artemis causes the lack of wind because he offended her. To appease the goddess, he has to sacrifice her daughter Iphigenia. So, in the moment before the prophecy, Agamemnon is committed to the achievement goal  $ag_{setSail}$ , motivated by the value of loving his country  $v_{Country}$ , and to the perform goal  $pg_{setSail}$  for the plan  $\pi_{setSail}$ , which consists in the departure from Aulis to Troy (Table 1).

**Table 1.** Agamemnon’s state before learning Calchas’s prophecy

Beliefs	$wind\text{-}intensity(low)$ $failed(\pi_{setSail})$ $inAulis(ships)$ $wind\text{-}intensity(high) \iff sailing(fleet)$
Desires	$ag_{setSail}(sailing(fleet))$ , $wind\text{-}intensity(low)$ , activated, $\pi_{setSail}$ , AG) $pg_{setSail}(succeeded(\pi_{setSail}), failed(\pi_{setSail}), activated, \pi_{setSail}, PG)$
Plans	$\pi_{setSail}(-, -, v_{Country}, ag_{setSail})$
$V_{atStake}$	$v_{Country}(inAulis(ships), 0.8, 0.5)$

In the Monitoring phase (Table 2) Agamemnon updates his beliefs. According to the prophecy, the ground formula  $not(alive(Iphigenia)) \iff wind\text{-}intensity(high)$  is added to the beliefs. The goals  $pg_{setSail}$  and  $ag_{setSail}$  are dropped because their failure condition ( $wind\text{-}intensity(low)$  and  $failed(\pi_{setSail})$ ) hold in the state of the world, while the set of values at stake  $V_{atStake}$  remains unaltered. So, the initial plan  $\pi_{setSail}$  is abandoned. In the Goal Adoption phase (Table 2), Agamemnon forms the achievement goal  $ag_{Country}$  to restore the value still at stake. In the Option Generation phase (Table 2) Agamemnon forms a new

plan  $\pi_{Country}$  to restore the value  $v_{Country}$  and satisfy the goal  $ag_{Country}$  according to his new beliefs. Agamemnon performs the anticipatory reasoning for new plans generated and he detects that the plan  $\pi_{Country}$  contains an action with an effects (i.e. ‘sacrifice Iphigenia’) that threatens a condition of value  $v_{Iphigenia}$  (preserving the life of her daughter Iphigenia). He forms a new value-dependent achievement goals  $ag_{Iphigenia}$  to re-establish  $v_{Iphigenia}$  in  $V$  and derives a new plan,  $\pi_{Iphigenia}$  (saving Iphigenia’s life) to achieve the new goal. Agamemnon performs a check to validate if and which plans are in conflict. At this point, a chiasmus is detected between plans  $\pi_{Country}$  and  $\pi_{Iphigenia}$ .

**Table 2.** Monitor - Goal Adoption - Option generation Phase 1

Beliefs	$wind-intensity(low)$ $inAulis(ships)$ $wind-intensity(high) \iff sailing(fleet)$ $not(alive(Iphigenia)) \iff wind-intensity(high)$
Desires	$ag_{Country}(sailing(fleet)), low-intensity(low), suspended, -, AG)$ $ag_{Iphigenia}(alive(Iphigenia), not(alive(Iphigenia)), suspended, -, AG)$
Plans	$\pi_{Country}(v_{Iphigenia}, v_{Country}, ag_{Country})$ $\pi_{Iphigenia}(v_{Country}, v_{Iphigenia}, ag_{Iphigenia})$
$V_{atStake}$	$v_{Country}(-, inAulis(ships), 0.8, 0)$ $v_{Iphigenia}(-, not(alive(Iphigenia)), 0.9, 0)$
$V_{Shared}$	(Menelaus, $v_{Country}$ , 1) with priority $(0.8 * 1) = 0.8$ (Ulysses, $v_{Country}$ , 0.5) with priority $(0.8 * 0.5) = 0.4$ (fleet, $v_{Country}$ , 0.3) with priority $(0.8 * 0.3) = 0.24$

In the Appraisal-based Deliberation phase (Table 3), due to the detected conflict, Agamemnon reasons about moral valence of plans and calculates the expected utility of plans according to the rules described in Section 3. We suppose that both plans contains two actions and a success probability of 0.5. So, for example, in the case of plan  $\pi_{Country}$  the pride emotion intensity is  $\sum_{v_{country}} r * step(\pi_i) * probOfSucc(\pi_i)$ ; the shame emotion intensity is  $\sum_{v_{Iphigenia}} r * (1 - l) * step(\pi_i) * probOfSucc(\pi_i)$  and the expected utility is equal to the difference between pride intensity and shame intensity. After the self-appraisal, Agamemnon takes into consideration the shared values to perform social appraisal (Table 3). The only value at stake shared with others agents is the value  $v_{Country}$ . The new priority for the shared value is calculated as the sum of the shared-priority (Table 2) with the original priority of the value (0.8). So the new priority for the value at stake  $v_{Country}$  is equal to  $(0.8 + 0.8 + 0.4 + 0.24) = 2.24$ . After the social appraisal, the highest expected utility is the expected utility of the plan  $\pi_{Country}$ . The achievement goal  $ag_{Country}$  becomes ‘activated’ and a perform goal  $pg_{Country}$  is formed to start executing the chosen plan. In the execution phase, Agamemnon executes the first action of the plan: he writes a letter to Iphigenia to tell her that she must join him.

**Table 3.** Appraisal-based Deliberation phase

	Self-appraisal $\pi_{Country}$	Self-appraisal $\pi_{Iphigenia}$
$Pride_{int}$	$0.8 * 2 * 0.5 = 0.8$	$0.9 * 2 * 0.5 = 0.9$
$Shame_{int}$	$(0.9 * (1-0)) * 2 * 0.5 = 0.9$	$(0.8 * (1-0)) * 2 * 0.5 = 0.8$
utility	$0.8 - 0.9 = -0.1$	$0.9 - 0.8 = 0.1$
	Social-appraisal $\pi_{Country}$	Social-appraisal $\pi_{Iphigenia}$
$Pride_{int}$	$2.24 * 2 * 0.5 = 2.24$	$0.9 * 2 * 0.5 = 0.9$
$Shame_{int}$	$(0.9 * (1-0)) * 2 * 0.5 = 0.9$	$2.24 * (1-0) * 2 * 0.5 = 0.8$
utility	$2.24 - 0.9 = 1.24$	$0.9 - 2.24 = -1.34$

In the subsequent cycle, Agamemnon monitors the world again and the appraisal restarts. For the sake of brevity, we focus on the Appraisal-based Deliberation phase. Agamemnon's mental state is depicted in Table 4. Following the story line, after Agamemnon has written the letter to his daughter, the shared values are removed from his belief base. In the goal adoption phase Agamemnon

**Table 4.** Monitoring - Goal adoption - Option Generation - Deliberation Phase 2

Beliefs	$wind-intensity(low)$ $inAulis(ships)$ $wind-intensity(high) \iff sailing(fleet)$ $not(alive(Iphigenia)) \iff wind-intensity(high)$ $coming(Iphigenia)$
Desires	$agCountry(sailing(fleet)), low-intensity(low), activated, -, AG)$ $pgCountry(succeeded(\pi_{Country}), failed(\pi_{Country}), activated, \pi_{Country}, PG)$ $agIphigenia(alive(Iphigenia), not(alive(Iphigenia)), suspended, -, AG)$
Plans	$\pi_{Country}(-, v_{Iphigenia}, v_{Country}, agCountry)$ $\pi_{Iphigenia}(-, v_{Country}, v_{Iphigenia}, agIphigenia)$
$V_{atStake}$	$v_{Country}(inAulis(ships), 0.8, 0)$ $v_{Iphigenia}(not(alive(Iphigenia)), 0.9, 0)$
$\pi_{Country}$	Self-appraisal
$Pride_{int}$	$0.8 * 2 * 0.9 = 1.44$
$Shame_{int}$	$(0.9 * (1-0)) * 2 * 0.9 = 1.62$
utility( $\pi_{Country}$ )	$1.44 - 1.62 = -0.18$
$\pi_{Iphigenia}$	Self-appraisal
$Pride_{int}$	$0.9 * 2 * 0.5 = 0.9$
$Shame_{int}$	$(0.8 * (1-0)) * 2 * 0.5 = 0.8$
utility( $\pi_{Iphigenia}$ )	$0.9 - 0.8 = 0.1$

is still committed to the goals  $agCountry$  and  $pgCountry$ , while the goal  $agIphigenia$  is still suspended. In the option generation phase there is no new value at stake, only the plan  $\pi_{Iphigenia}$  may contain a different course of action to re-establish the value still at stake, because of the new beliefs (Table 4). In the Appraisal-based Deliberation phase, Agamemnon re-calculates the utility of plans. This time, the success probability of the plan  $\pi_{Country}$  is equal to 0.9 (he has writ-

ten the letter and Iphigenia is bound to arrive). We don't report the social appraisal re-calculation because no values are shared with the society at this time (Table 4). Agamemnon chooses the best plan  $\pi_{Iphigenia}$ . Consequently, the achievement goal  $ag_{Country}$  becomes suspended, the perform goal  $pg_{Country}$  and the plan  $\pi_{Country}$  are dropped, while  $ag_{Iphigenia}$  is activated and a new perform goal is formed to execute the new plan. Then, in the Execution phase, Agamemnon starts to execute the plan  $\pi_{Iphigenia}$ . According to the story line, he feels remorse for what he has done, and writes a new letter to his daughter saying her not to come to Aulis.<sup>1</sup>

## 6 Conclusion and Future Work

In this paper, we proposed an agent model that accounts for the emotional range triggered by the occurrence of a moral dilemma in stories. We introduced the notion of values to define the moral appraisal of actions and illustrated the model on a literary example of moral dilemma. The aim is to create artificial characters whose emotional appraisal is affected by moral values, and shared values in particular. Being an operational model, it lends itself to implementation with agent frameworks and languages, that we are currently conducting.

We want to extend our model to consider not only value-dependent goals but also standard goals, in order to investigate the interplay of value-dependent goal formation and standard deliberation, and generate a wider range of emotional states, that account for the social component in a more accurate way. Last, we want to test the emotional states generated by the model on the involvement of the user and her perception of the character's behavior, by testing if the moral values of the user have an impact in the mechanism of sympathy and empathy.

## References

1. R. Aylett, M. Vala, P. Sequeira, and A. Paiva. Fearnot!—an emergent narrative approach to virtual dramas for anti-bullying education. *LNCS*, 4871:202, 2007.
2. H. Barber and D. Kudenko. Generation of dilemma-based interactive narratives with a changeable story goal. In *Proc. of 2nd INTETAIN*, Brussels, 2008.
3. Christian Becker-Asano. *WASABI: Affect Simulation for Agents with Believable Interactivity*. PhD thesis, Faculty of Technology, University of Bielefeld, 2008. IOS Press (DISKI 319).
4. Rafael Bordini and Jomi Hübner. Bdi agent programming in agentspeak using jason. In Francesca Toni and Paolo Torroni, editors, *Computational Logic in Multi-Agent Systems*, volume 3900 of *Lecture Notes in Computer Science*, pages 143–164. Springer Berlin / Heidelberg, 2006.
5. Michael Bratman. *Intentions, Plans, and Practical Reason*. Harvard University Press, 1987.

---

<sup>1</sup> It is interesting to note that the story says that Menelaus prevents Agamemnon from sending his letter. When Iphigenia arrives to Aulis, she faces exactly the same dilemma as her father. But in this case, since personal values are different, the conclusion is also different: she decides to sacrifice herself for the sake of her country.

6. J. Bruner. The narrative construction of reality. *Critical Inquiry*, 18(1):1–21, 1991.
7. L. Lesmo C. Battaglino, R. Damiano. Moral appraisal and emotions. In *Workshop EEA - Emotional and Empathic Agents, AAMAS*, 2012.
8. Noel Carroll. Art and mood: preliminary notes and conjectures. *The monist*, 86(4):521–555, 2003.
9. P. Cohen and H. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42 (2-3):213–261, 1990.
10. R. Damiano and V. Lombardo. An Architecture for Directing Value-Driven Artificial Characters. *Agents for Games and Simulations II: Trends in Techniques, Concepts and Design*, pages 76–90, 2010.
11. R. Damiano and V. Lombardo. An Architecture for Directing Value-Driven Artificial Characters. *Agents for Games and Simulations II: Trends in Techniques, Concepts and Design*, pages 76–90, 2011.
12. L. Egri. *The Art of Dramatic Writing*. Simon and Schuster, New York, 1946.
13. Magy Seif El-Nasr, John Yen, and Thomas R. Ioerger. Flame-fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems*, 3:219–257, September 2000.
14. Clark D. Elliott. *The affective reasoner: a process model of emotions in a multi-agent system*. PhD thesis, Northwestern University, Evanston, IL, USA, 1992.
15. Susan L. Feagin. On Noel Carrol on narrative closure. *Philosophical Studies*, (135):17–25, 2007.
16. Patrick Gebhard. Alma: a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems, AAMAS '05*, pages 29–36, New York, NY, USA, 2005. ACM.
17. Richard S. Lazarus. *Emotion and Adaptation*. Oxford University Press, USA, August 1991.
18. S. Marsella and J. Gratch. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, March 2009.
19. Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, July 1988.
20. F. Peinado, M. Cavazza, and D. Pizzi. Revisiting Character-based Affective Storytelling under a Narrative BDI Framework. In *Proc. of ICIDIS08*, Erfurt, Germany, 2008.
21. A. Pokahr, L. Braubach, and W. Lamersdorf. Jadex: a BDI Reasoning Engine. *Multiagent Systems, Artificial Societies and Simulated Organizations*, 15:149, 2005.
22. G. Polti. *Les trente-six situations dramatiques*. Mercure de France, Paris, 1895.
23. Julie Porteous, Marc Cavazza, and Fred Charles. Narrative generation through characters’ point of view. In *AAMAS*, pages 1297–1304, 2010.
24. W. Scott Reilly and Joseph Bates. Building emotional agents, 1992.
25. K. R. Scherer. The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology*, 73:902–922, 1997.
26. B. van Fraassen. Values and the heart’s command. *Journal of Philosophy*, 70(1):5–19, 1973.
27. M.B. van Riemsdijk, M. Dastani, and M. Winikoff. Goals in Agent Systems: A Unifying Framework. In *Proceedings of AAMAS’08*, 2008.
28. Stanley D. Williams. *The Moral Premise*. Michael Wiese Productions, 2006.